



Lending Club Case Study: Pre-Assignment Session

1

Case Study: Exploratory Data Analysis

Lending Club Data Set

Group Facilitator : Taruj Bhattacharya

Group Member : Nikhil Shinde

March 2022

What is Lending Club?

Lending Club is a marketplace for personal loans that matches borrowers who are seeking a loan with investors looking to lend money and make a return.

How Lending Club Works



Borrowers apply for loans.
Investors open an account.



Borrowers get funded.
Investors build a portfolio.



Borrowers repay automatically.
Investors earn & reinvest.

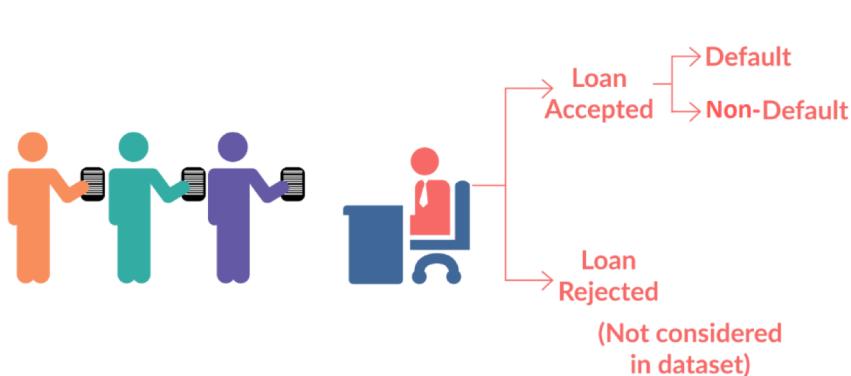
When the company receives a loan application, the company has to make a decision for loan approval based on the applicant's profile. Two types of risks are associated with the bank's decision:

- If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company
- If the applicant is not likely to repay the loan, i.e. he/she is likely to default, then approving the loan may lead to a financial loss for the company

If one is able to identify these risky loan applicants, then such loans can be reduced thereby cutting down the amount of credit loss. Identification of such applicants using EDA is the aim of this case study.

In other words, the company wants to understand the **driving factors (or driver variables)** behind loan default, i.e. the variables which are strong indicators of default. The company can utilize this knowledge for its portfolio and risk assessment.





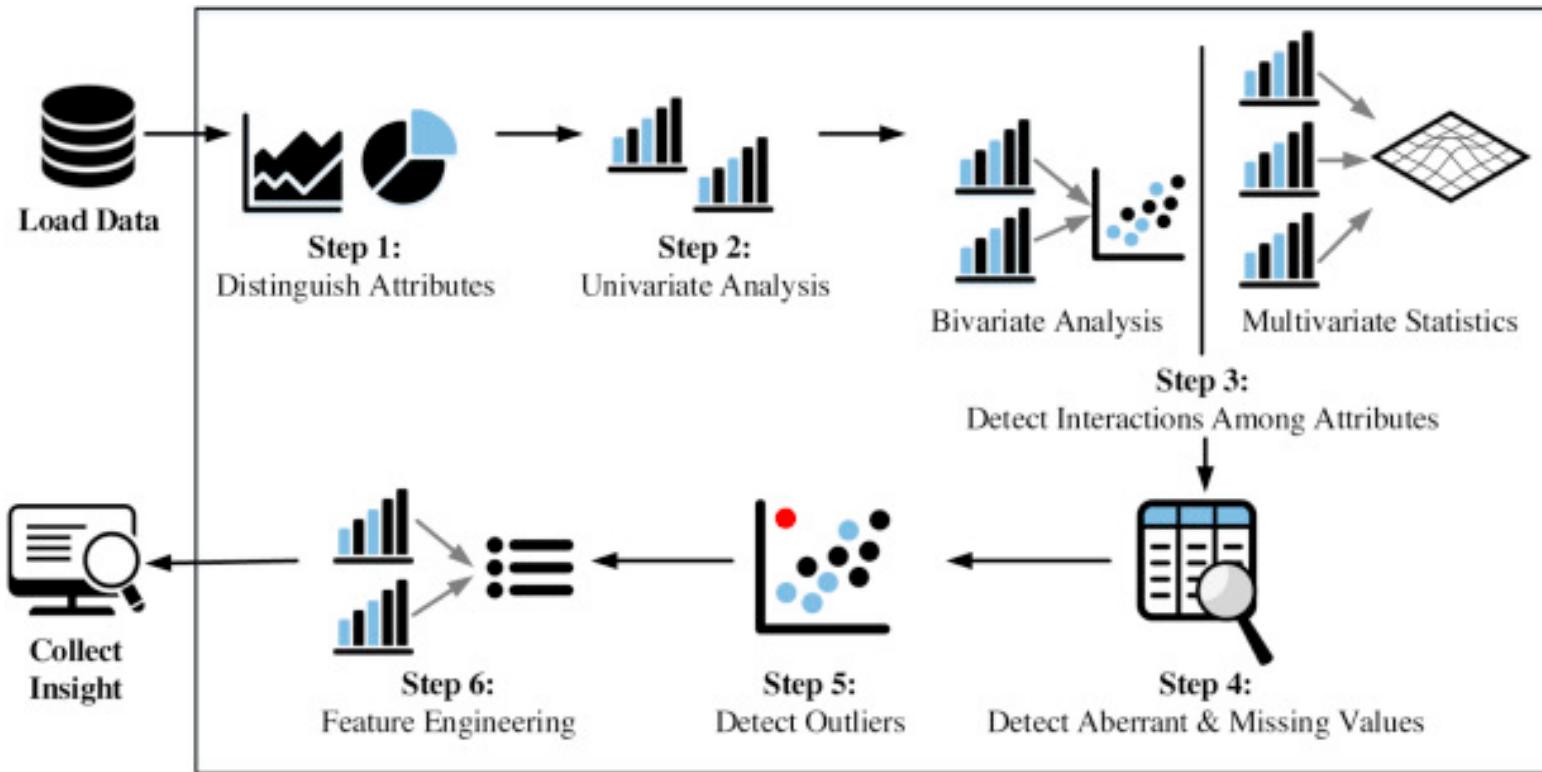
Fully paid: Applicant has fully paid the loan (the principal and the interest rate)

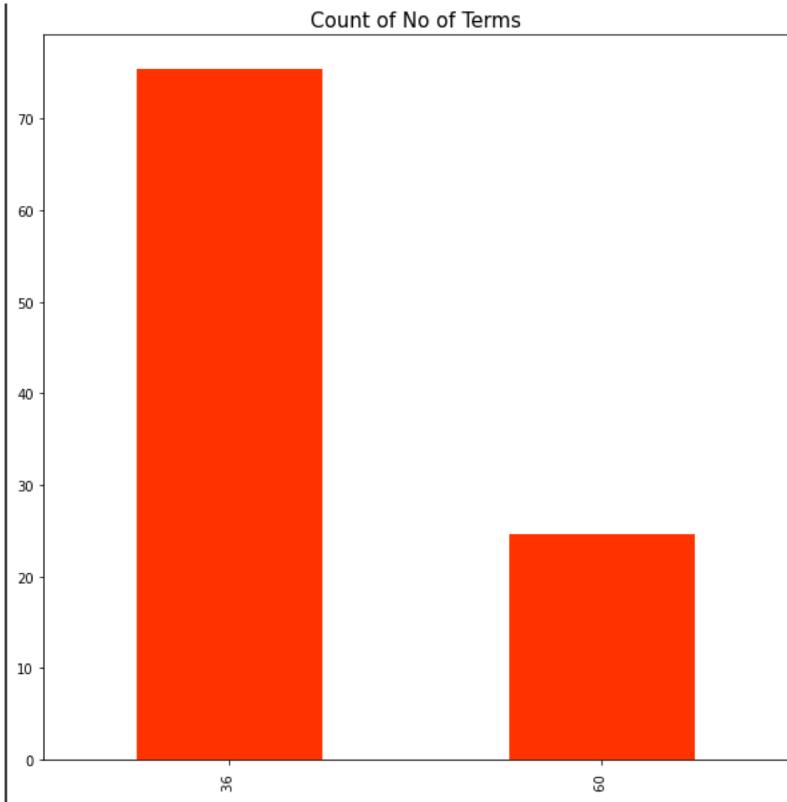
Current: Applicant is in the process of paying the instalments, i.e. the tenure of the loan is not yet completed. These candidates are not labelled as 'defaulted'.

Charged-off: Applicant has not paid the instalments in due time for a long period of time, i.e. he/she has defaulted on the loan

Fundamental steps of EDA process

upGrad



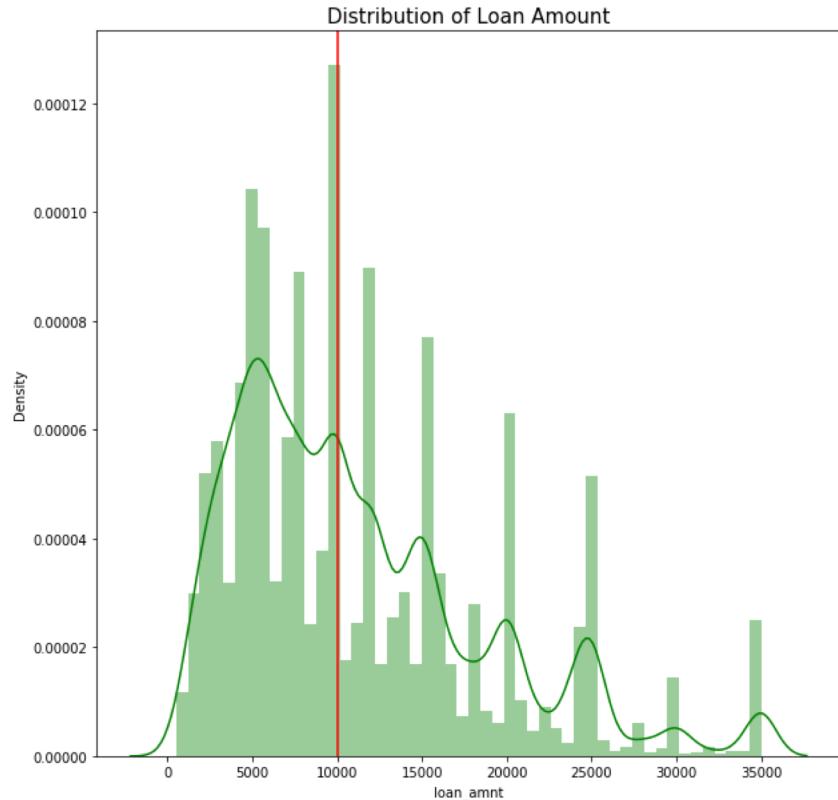


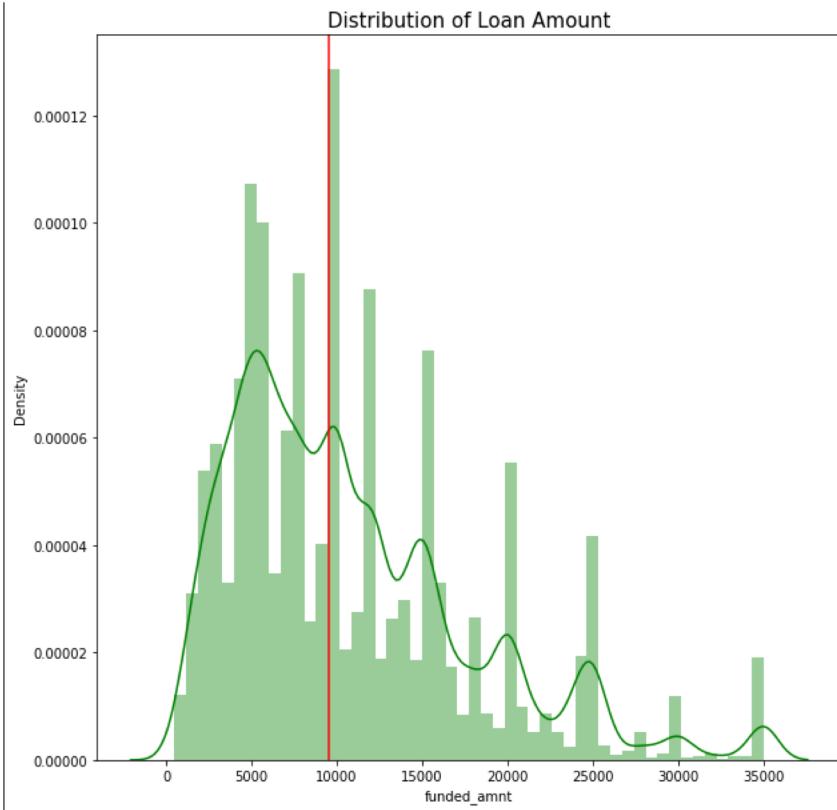
Term – (The Duration of the loan)

From the figure we can see most of the loans are given for the term of 36 Months (3 Years) or 60 Months (5 Years)

loan_amnt – Duration of the Loan

- Median loan amount requested is 10000
- Skewness is high in loan_amnt since more people have requested for loans above 10000 which is median

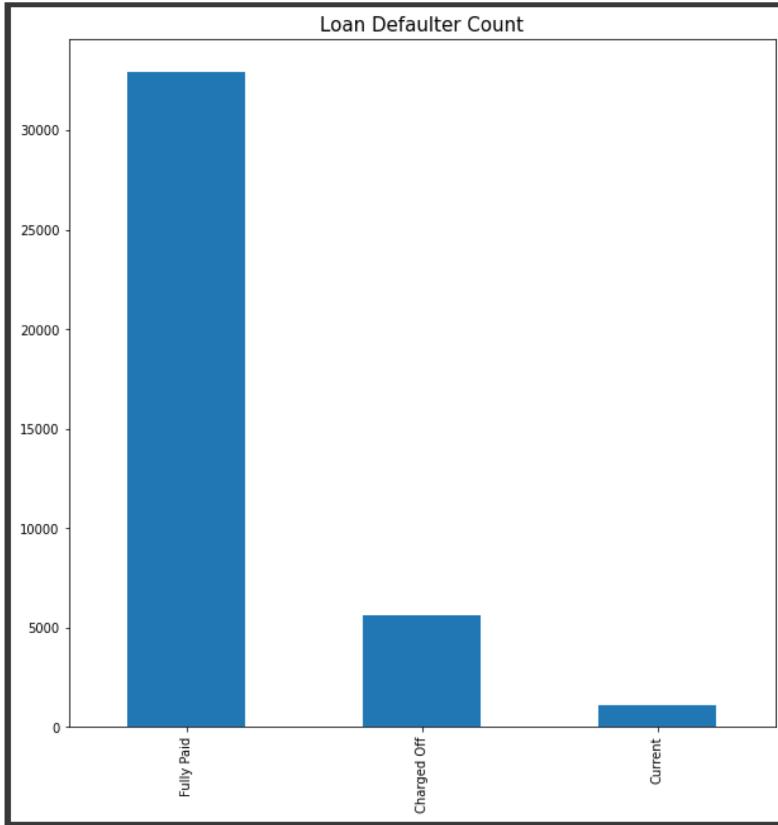




Loan Requested & Loan Funded

The figure clearly shows that the loan_amount requested and loan_amount funded has difference and is apparent for the median of \$400

Funded amount is Right Skewed



Loan Status - Numerical Distribution

Most loans are Fully Paid

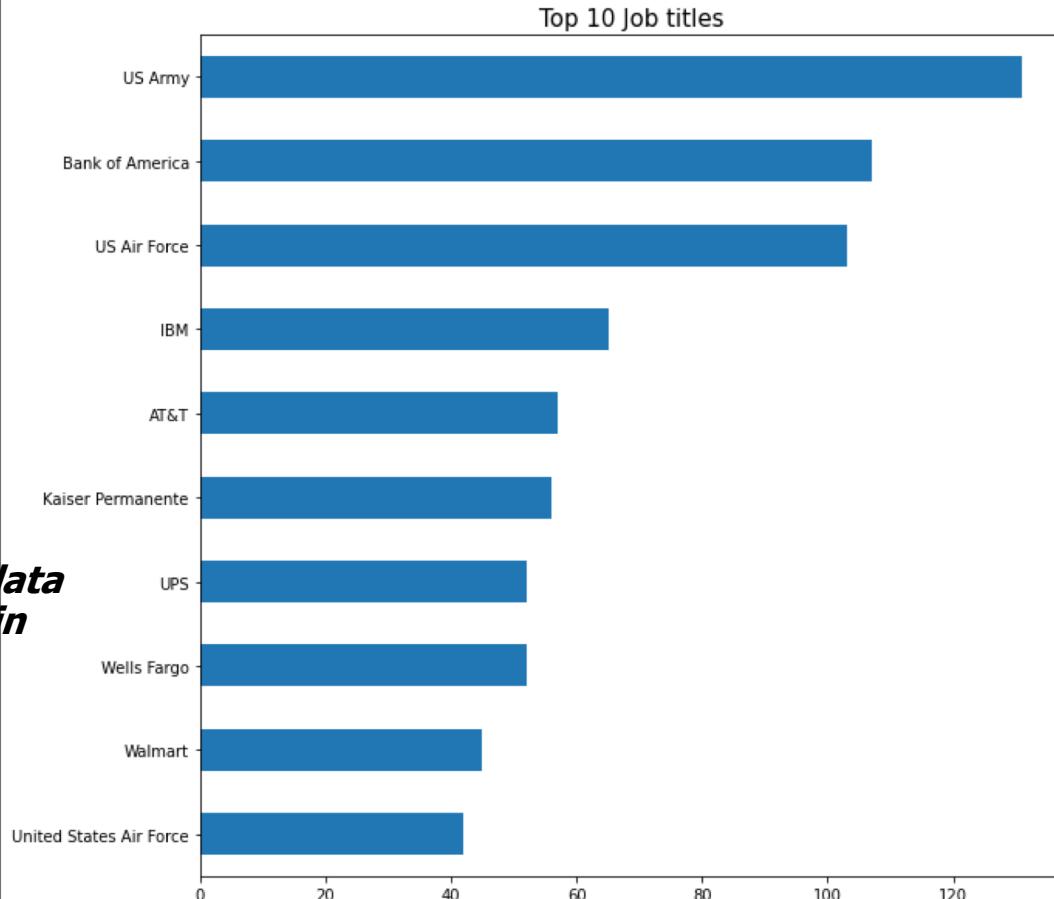
Current Loans are being Paid

Charged Off – The loan (borrowers) that have defaulted

Loans by Job Title

US Army, Bank of America & IBM Employees are the most frequent borrowers

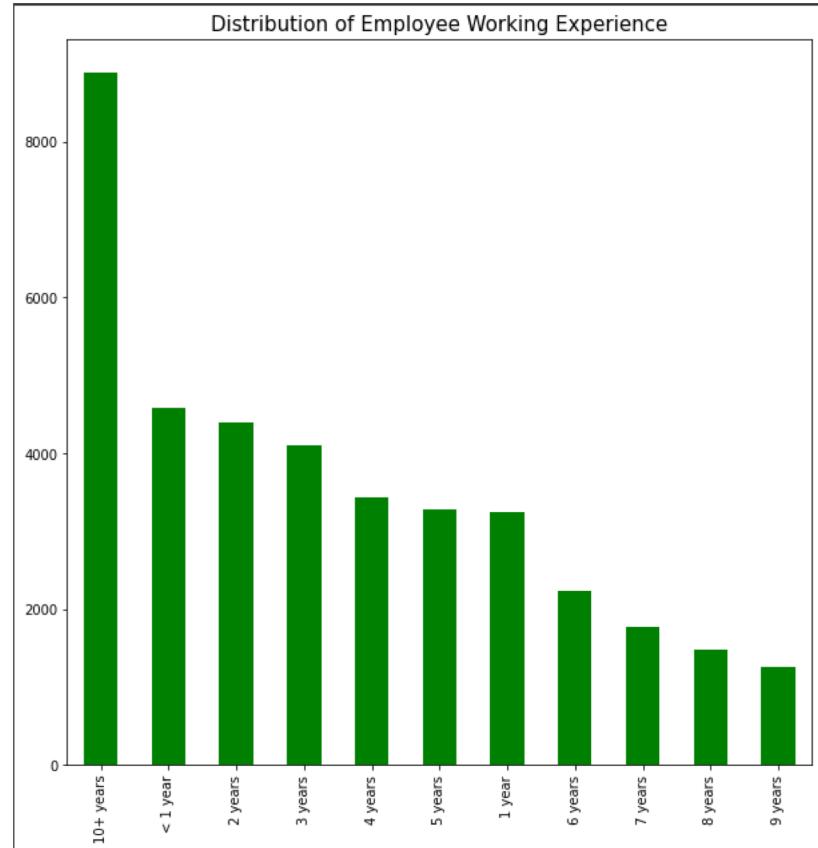
The figure and the data discrepancy in the data where the USAF and US air force should be in same category

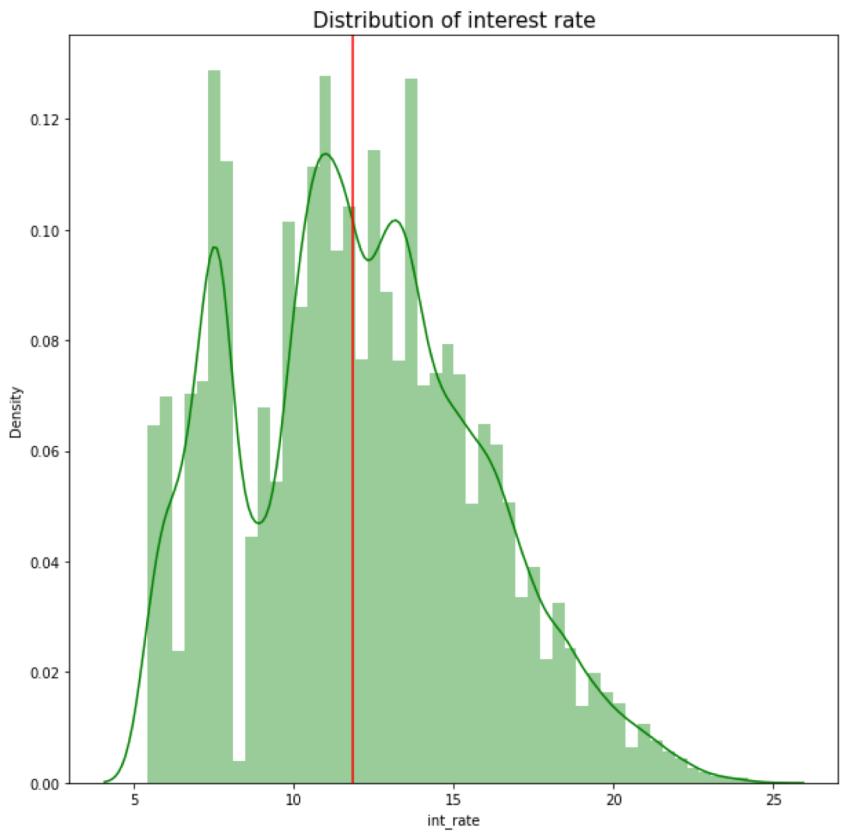


We can see that the data shows that people with 10+ years of experience takes MOST loans

And people with Experience 9 years take dramatically Least Loans

Points to that people after 10 years lot of loans compared to people with less experience than 10 years





Interest Rate Distribution

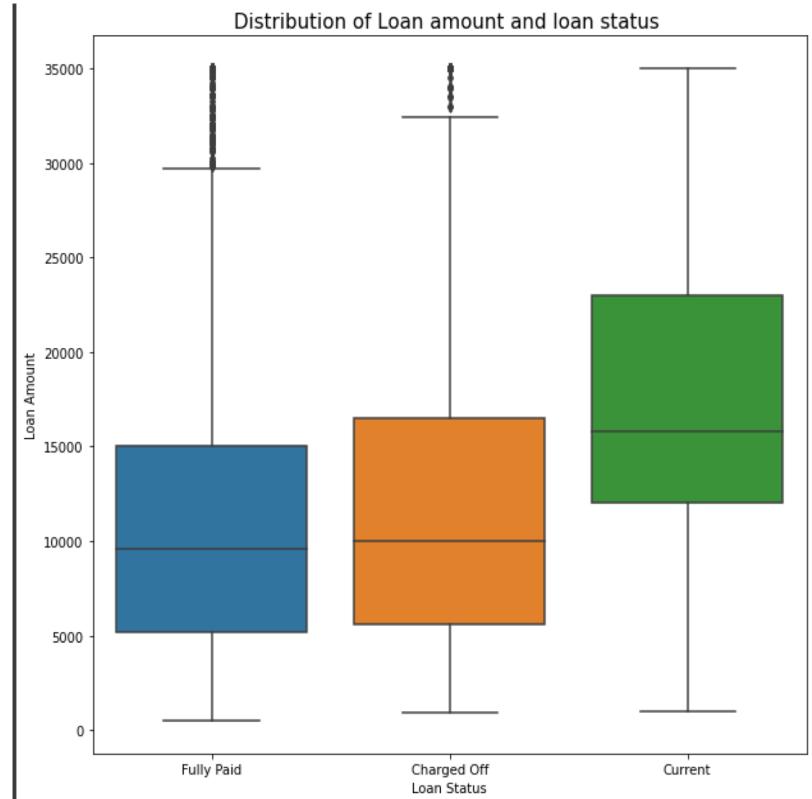
The median interest rates are 11.86 % and the data is almost normally distributed

LOAN AMOUNT VS LOAN STATUS

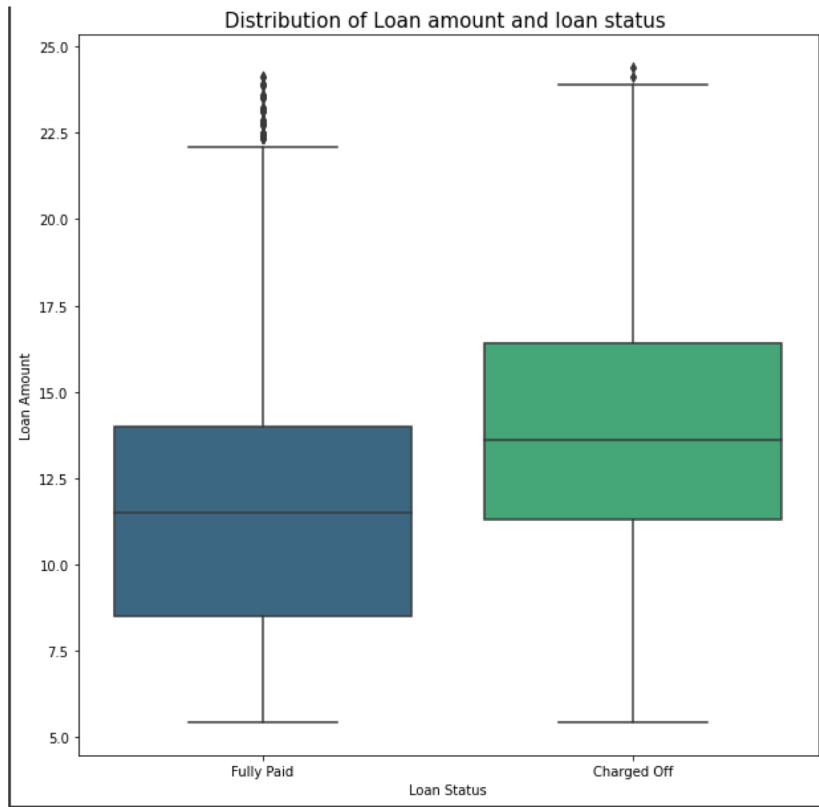
We can see that the data loan status has 3 categories and the current category doesn't seem to helpful in Prediction so we should remove it

The loans that are charged off are usually in the range of 5k to 16k and with some outliers

People tend to pay it more often than charging it off



Lending Club: EDA Case Study



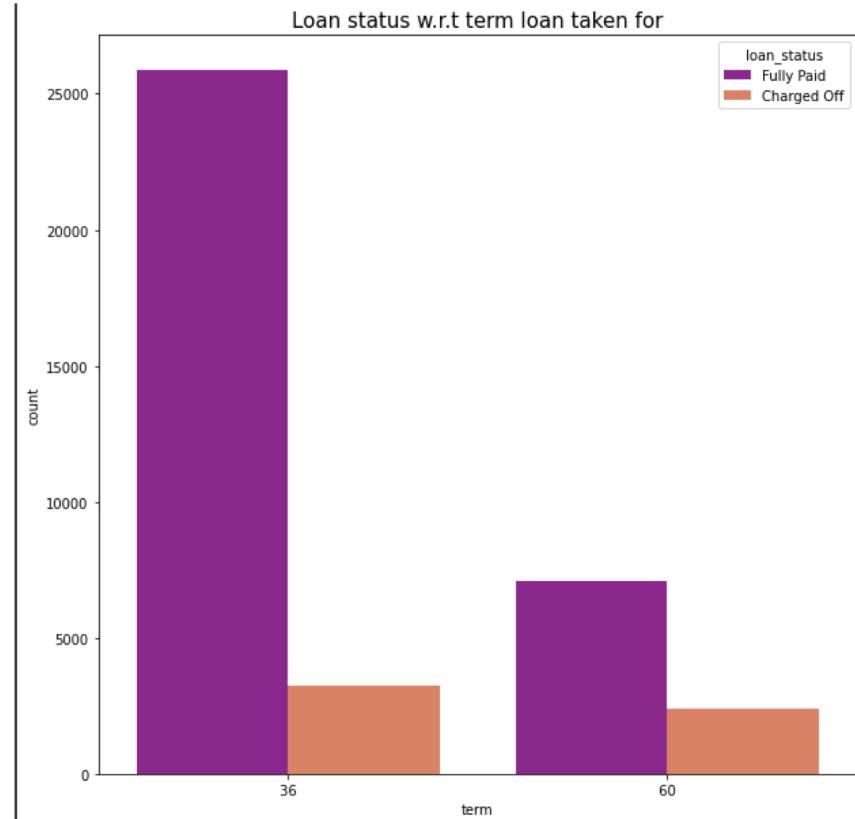
Loan Status vs Interest Rates

The data shows that the loans Charged off are the loan which are higher than 11% and most of the charged off are because of higher interest Rates

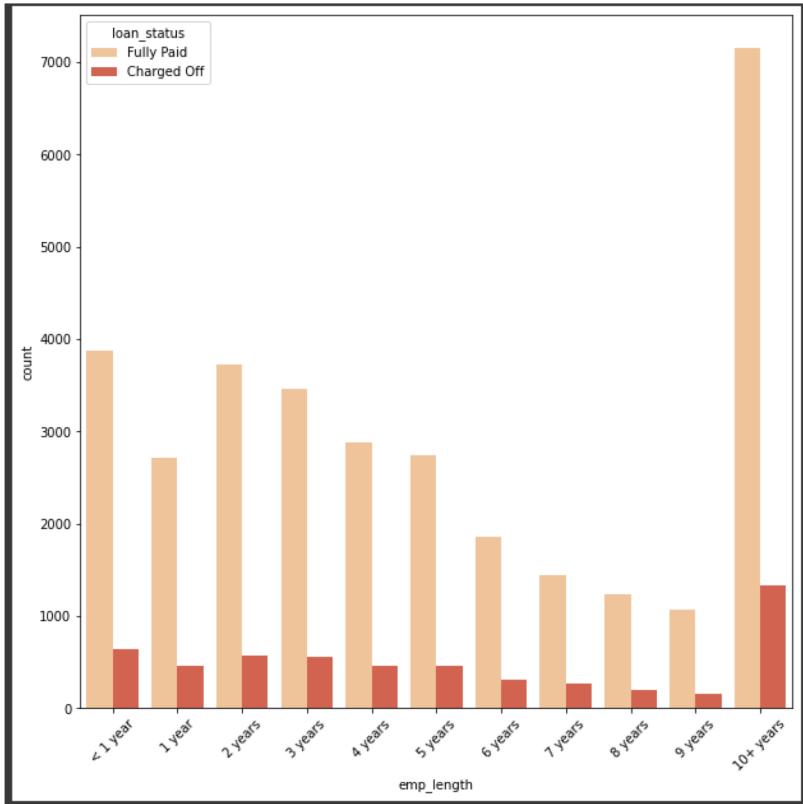
Loan Status wrt to Loan Duration

We can see that the loans are charged off more often in 36 months period while previously we saw that because of the interest rate the loan were charged off

We can form an assumption that the loan with less term and high interest rates are waived off **usually**



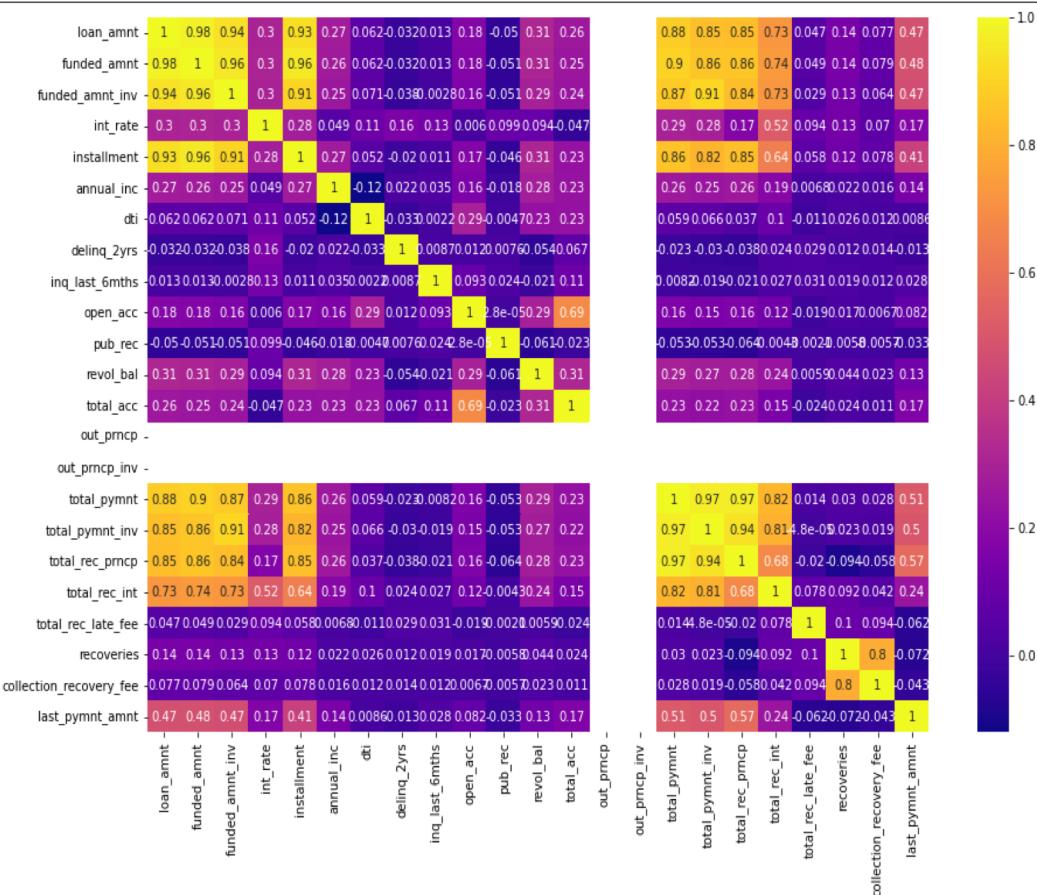
Lending Club: EDA Case Study



Most charge off if the employee is having experience of more than 10 years

Also the employee who seem to have less than 1 year experience seem to take lot of loans and charge them off

Lending Club: EDA Case Study



High Correlation Between
• loan_amnt to installment

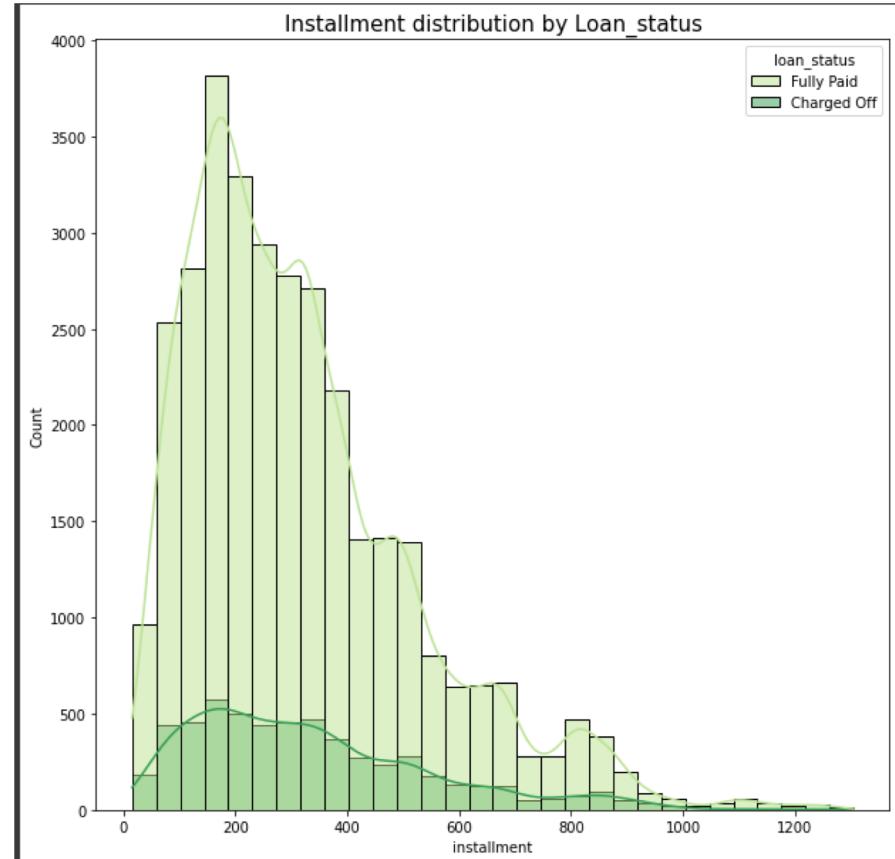
• collection_recovery to recovery

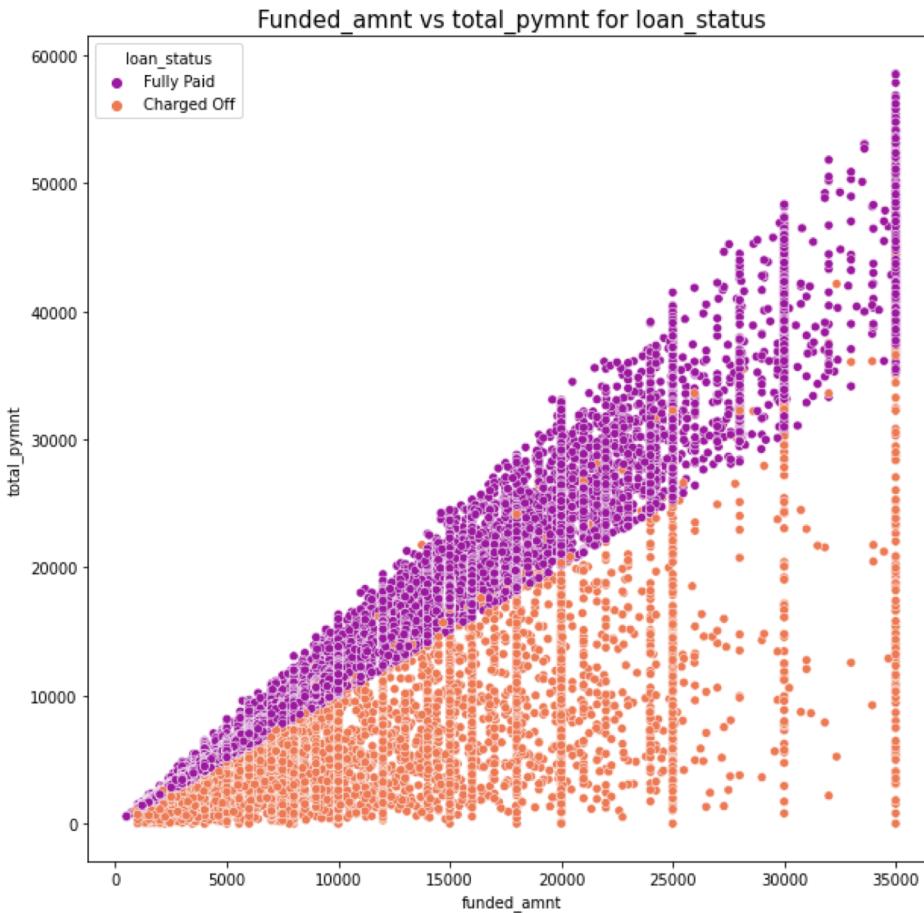
• Last recorded interest to funded amount

Lending Club: EDA Case Study

We can see that the most of the distribution of the installment lies between 0 to 600 who have charged off there loans

While we can see that most of people fully paid there loans lies between 0 to 800

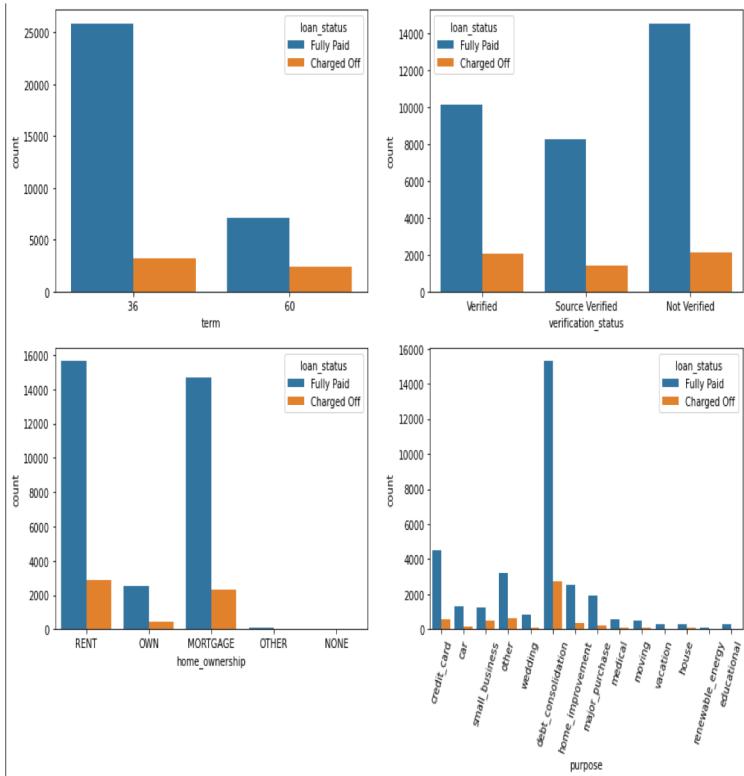




Funded Amount Vs Total Payment

We can see that the data is creating a funnel like shape

When the funded amount is increasing the total_pymnt is almost constant for the charged off that is 0 and which can be assumed that they havent paid any installments



Critical Insights

- People who are taking the loans for 60 terms have lower defaulter rate than who takes it for lower terms**
- The verification status shows that the most of people who are source verified tent default loans very less than while when the source isn't verified the person tend to default the loan**
- If the person having the own house its seems very rare that he would default the loan and the rented houses tend to default the loans more often**
- when a person takes the loan for debt consolidation there is more risk the person will default the loans**



Thank You!

<https://towardsdatascience.com/exploratory-data-analysis-8fc1cb20fd15>

<https://aditya-bhattacharya.net/2021/09/18/daily-dose-of-data-science-day-5-attractive-data-visualizations-in-python-using-matplotlib-styles-and-plotly/>